ON THE RANK-ONE APPROXIMATION OF SYMMETRIC TENSORS

MICHAEL J. O'HARA*

Abstract. The problem of symmetric rank-one approximation of symmetric tensors is important in Independent Components Analysis, also known as Blind Source Separation, as well as polynomial optimization. We analyze the symmetric rank-one approximation problem for symmetric tensors and derive several perturbation results. Given a symmetric rank-one tensor obscured by noise, we provide bounds on the accuracy of the best symmetric rank-one approximation for recovering the original rank-one structure, and we show that any eigenvector with sufficiently large eigenvalue is related to the rank-one structure as well. Further, we show that for high-dimensional symmetric approximately-rank-one tensors, the generalized Rayleigh quotient is mostly close to zero, so the best symmetric rank-one approximation corresponds to a prominent global extreme value. We show that each iteration of the Shifted Symmetric Higher Order Power Method (SS-HOPM), when applied to a rank-one symmetric tensor, moves towards the principal eigenvector for any input and shift parameter, under mild conditions. Finally, we explore the best choice of shift parameter for SS-HOPM to recover the principal eigenvector. We show that SS-HOPM is guaranteed to converge to an eigenvector of an approximately rank-one even-mode tensor for a wider choice of shift parameter than it is for a general symmetric tensor. We also show that the principal eigenvector is a stable fixed point of the SS-HOPM iteration for a wide range of shift parameters; together with a numerical experiment, these results lead to a non-obvious recommendation for shift parameter for the symmetric rank-one approximation problem.

Key words. symmetric rank-one approximation, symmetric tensors, tensors, higher-order power method, shifted higher-order power method, tensor eigenvalues, Z-eigenpairs, l_2 eigenpairs, blind source separation, independent components analysis

AMS subject classifications. 15A69

1. Introduction. The symmetric rank-one approximation of a symmetric tensor has at least two important applications. One is Independent Components Analysis, known in signals processing as Blind Source Separation [1, 6]. First, we recall classical Principal Components Analysis (PCA). PCA identifies a basis for a set of random variables that diagonalizes the covariance matrix, in other words a basis where the random variables are uncorrelated. This is necessary but not sufficient for independence. A stronger test for independence is to check whether the off-super-diagonal elements of the four-way cumulant tensor, a symmetric tensor defined from the fourth-order statistical moments, are zero. A linear transformation that achieves this can be identified by writing the tensor as a sum of symmetric rank-one terms; one approach uses successive symmetric rank-one approximations [11].

Another important application of the symmetric rank-one approximation of symmetric tensors is in the optimization of a general homogeneous polynomial over unit length vectors, i.e. the *unit sphere* [8]. For instance, the symmetric rank-one variant of the "Time Varying Covariance Approximation 2" (TVCA2) problem [10] can be written

$$\max_{x:\ ||x||=1} \sum_{t=1}^{T} (x^T A_t x)^2 \ , \tag{1.1}$$

where $\{A_t : t = 1 \dots T\}$ are a given set of covariance matrices, and the vector norm is the 2-norm (as are all subsequent norms unless otherwise indicated). The argument of

*Mailbox L-363, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore CA 94550 (ohara7@llnl.gov).

(1.1) is a degree-4 homogeneous polynomial, and so as we will see the TVCA2 problem can be represented as the best symmetric rank-one approximation of a symmetric tensor.

Some things are known about the symmetric rank-one approximation problem. The best symmetric rank-one approximation in the Frobenius norm corresponds to the principal tensor eigenvector, and also the global extreme value of the generalized Rayleigh quotient [3, 5]. It is not clear that these facts help us solve the symmetric rank-one approximation problem, because tensor computations are generally notoriously difficult. For instance it is known [2] that (asymmetric) rank-one approximation of a general mode-3 tensor is NP-complete. However, there is an algorithm, the Symmetric Shifted Higher Order Power Method (SS-HOPM) [4], that is guaranteed to find symmetric tensor eigenvectors.

We address several questions pertaining to the rank-one approximation of symmetric tensors. In Section 3, we address the structure of approximately-rank-one symmetric tensors. A symmetric rank-one tensor obscured with noise has a best symmetric rank-one approximation that may not be the same as the original unperturbed tensor; how close is it? Is only the principal eigenvector related to the rank-one structure? For a given symmetric approximately-rank-one tensor, how well-separated is the principal eigenvalue from the spurious eigenvalues? In Section 4, we consider the application of SS-HOPM to approximately-rank-one symmetric tensors. How is the convergence of SS-HOPM affected by the approximately-rank-one structure? When does SS-HOPM find the principal eigenvector? We employ a perturbation approach to prove six theorems that provide insight all these questions.

2. Background and notation. A *tensor* is a multi-dimensional array of numbers. The number of *modes* of the tensor, m, is the number of indices required to specify entries; a mode-2 tensor is a matrix. The range of permissible index values $(n_1, \dots n_m)$ are the *dimensions* of the tensor; if all the dimensions are the same, as with symmetric tensors, we simply write n. A *symmetric tensor* has entries that are invariant under permutation of indices. For instance, for a mode-3 symmetric tensor $\mathcal{A}$, we have $\mathcal{A}_{123} = \mathcal{A}_{231}$. In this paper, tensors will be represented with script capital letters, matrices with capital letters, vectors with lower-case letters, and real numbers with lowercase Greek letters. Integers such as indices, dimensions, etc. will also be lowercase letters (e.g. $m, n, i \dots$).

A *symmetric rank-one tensor* is the outer product of a vector with itself, which we denote using the $\otimes$ operator. For instance, given the vector a, we can construct a symmetric rank-one tensor

$$(\underbrace{a \otimes a \otimes \cdots a}_{m \text{ times}})_{i_1 i_2 \dots i_m} \equiv (a^{\otimes m})_{i_1 i_2 \dots i_m} = a_{i_1} a_{i_2} \dots a_{i_m} \ . \tag{2.1}$$

The *rank* of a symmetric tensor $\mathcal{A}$ is the fewest number of symmetric rank-one terms whose sum is $\mathcal{A}$.

Generally, the $m - r$ product of the m-mode tensor $\mathcal{A}$ with the vector x is the r-mode tensor defined

$$(\mathcal{A}x^{m-r})_{i_1 \dots i_r} = \sum_{i_{r+1}, \dots i_m = 1}^{n} \mathcal{A}_{i_1 \dots i_m} x_{i_{r+1}} \dots x_{i_m} \ . \tag{2.2}$$

The special case $r = 0$ evaluates to a scalar and, under the constraint $||x|| = 1$, is called the *generalized Rayleigh quotient* [12]. Interestingly, any degree-m homogenous

polynomial, such as (1.1), can be written as $\mathcal{A}x^m$ for some symmetric tensor $\mathcal{A}$ and indeterminate x. In a miracle of notation, the derivatives are conveniently represented. The gradient may be written [4]

$$\nabla \mathcal{A}x^m = m\mathcal{A}x^{m-1} , \tag{2.3}$$

and the Hessian may be written [4]

$$\nabla^2 \mathcal{A}x^m = m(m-1)\mathcal{A}x^{m-2} . \tag{2.4}$$

The problem of maximizing the generalized Rayleigh quotient has the following Lagrangian:

$$\mathcal{L}(x, \mu) = \mathcal{A}x^m + \mu(x^T x - 1) , \tag{2.5}$$

where μ is the Lagrange multiplier. Using (2.3), we see the critical points of (2.5) satisfy the following symmetric tensor eigenproblem

$$\mathcal{A}x^{m-1} = \lambda x . \tag{2.6}$$

Solutions to (2.6) with $||x|| = 1$ are called *Z eigenvalues and eigenvectors* [7] to distinguish (2.6) from other tensor eigenvector problems, but here we will simply call them eigenvectors and eigenvalues. Together, we call an eigenvector and eigenvalue an *eigenpair*. The *principal eigenvector/value/pair* is that corresponding to the largest-magnitude eigenvalue, which may not be unique. For instance, if (x, λ) is an eigenpair, then if m is even so is $(-x, \lambda)$, otherwise if m is odd then so is $(-x, -\lambda)$ [4]. We will restrict our attention to real solutions to (2.6).

We note that symmetric tensor eigenvectors do not share all the properties of symmetric matrix eigenvectors, for instance they may not be orthogonal. Z eigenvectors are not scale-invariant so limiting our discussion to normalized eigenvectors is important. Finally, we note that because of the relationship between (2.6) and (2.5), the principal eigenvector corresponds to the extreme value of the generalized Rayleigh quotient, and the outer product of the principal eigenvector with itself, times the principal eigenvalue, is the best symmetric rank-one approximation of $\mathcal{A}$ in the Frobenius norm [3, 5].

The Shifted Symmetric Higher Order Power Method (SS-HOPM) [4], for a symmetric tensor $\mathcal{A}$, consists of the iteration

$$x_{k+1} = \frac{\mathcal{A}x_k^{m-1} + \alpha x_k}{||\mathcal{A}x_k^{m-1} + \alpha x_k||} , \tag{2.7}$$

where α is a scalar *shift parameter*. An eigenvector x is a stable fixed point of this iteration provided that the Hessian matrix for (2.7) is positive semidefinite at x. That condition is known [4] to be equivalent, for all $y \perp x$ and $||y|| = 1$, to

$$\left| \frac{(m-1)y^T \mathcal{A}x^{m-2} y + \alpha}{\lambda + \alpha} \right| < 1 . \tag{2.8}$$

It is known [4] that eigenpairs corresponding to local maxima of the generalized Rayleigh quotient (called *negative stable* eigenvectors) are stable fixed points of SS-HOPM provided $\alpha > \beta(\mathcal{A})$, where

$$\beta(\mathcal{A}) = (m-1) \max_{x:\, ||x||=1} \rho(\mathcal{A}x^{m-2}) , \tag{2.9}$$

and ρ returns the spectral radius of a matrix. Further, it is known [4] that if $\alpha > \beta(\mathcal{A})$, the SS-HOPM iteration monotonically increases the generalized Rayleigh quotient and converges to a tensor eigenvector. It is not clear how to compute $\beta(\mathcal{A})$, but we have the crude bound [4]

$$\beta(\mathcal{A}) \leq \hat{\beta}(\mathcal{A}) = (m-1) \sum_{i_1 i_2 \ldots i_m} |\mathcal{A}_{i_1 i_2 \ldots i_m}| \ . \tag{2.10}$$

The following three properties are useful.

LEMMA 1. *For any n dimensional vectors a and x, nonnegative integers m and $0 \leq r \leq m$, the following holds:*

$$(a^{\otimes m}) x^{m-r} = (a^T x)^{m-r} a^{\otimes r} \ . \tag{2.11}$$

Proof. We use (2.2) and (2.1):

$$\begin{aligned}
\left((a^{\otimes m}) x^{m-r}\right)_{i_1 \ldots i_r} &= \sum_{i_{r+1} \ldots i_m = 1}^{n} (a^{\otimes m})_{i_1, \ldots i_m} x_{i_{r+1}} \ldots x_{i_m} && (2.12)\\
&= \sum_{i_{r+1}, \ldots i_m = 1}^{n} a_{i_1} a_{i_2} \ldots a_{i_m} x_{i_{r+1}} \ldots x_{i_m} && (2.13)\\
&= a_{i_1} a_{i_2} \ldots a_{i_r} \left(\sum_{i_{r+1}=1}^{n} a_{i_{r+1}} x_{i_{r+1}} \right) \ldots \left(\sum_{i_m = 1}^{n} a_{i_m} x_{i_m} \right) && (2.14)\\
&= (a^T x)^{m-r} (a^{\otimes r})_{i_1 \ldots i_r} \ . && (2.15)
\end{aligned}$$

□

LEMMA 2 (Kolda and Mayo 2011 [4]). *For any m-mode symmetric tensor $\mathcal{A}$, and any unit-length vector x,*

$$|\mathcal{A} x^m| < \frac{\beta(\mathcal{A})}{m-1} \ . \tag{2.16}$$

LEMMA 3. *For any m-mode symmetric tensors $\mathcal{A}$ and $\mathcal{B}$, any vector x, and nonnegative integers m and $0 \leq r \leq m$, we have*

$$(\mathcal{A} + \mathcal{B}) x^{m-r} = \mathcal{A} x^{m-r} + \mathcal{B} x^{m-r} \ . \tag{2.17}$$

Lemma (3) follows directly from the definition of tensor-vector multiplication in (2.2).

3. Structure of approximately-rank-one symmetric tensors. Define

$$\mathcal{A} = \lambda \cdot a^{\otimes m} + \mathcal{E} \ , \tag{3.1}$$

where a is a unit-length n dimensional vector and $\mathcal{E}$ is a symmetric tensor representing noise. Clearly if $\mathcal{E} = 0$, then (a, λ) is a principal eigenpair, and all unrelated eigenvalues are zero. Now let us consider how close is (a, λ) to a principal eigenpair when $\mathcal{E} \neq 0$.

THEOREM 1. *Let $\mathcal{A}$ be defined by (3.1). Then a principal eigenvalue λ_p obeys*

$$|\lambda| - \frac{\beta(\mathcal{E})}{m-1} \leq |\lambda_p| \leq |\lambda| + \frac{\beta(\mathcal{E})}{m-1} \ , \tag{3.2}$$

and the angle θ between a and the corresponding principal eigenvector x_p is bounded by

$$|\cos^m \theta| \geq 1 - \frac{2\beta(\mathcal{E})}{|\lambda|\,(m-1)} \ . \tag{3.3}$$

Proof. Since (x_p, λ_p) are a tensor eigenpair, and $||x_p|| = 1$, we have

$$\mathcal{A}x_p^m = \lambda_p \ . \tag{3.4}$$

Using Lemma 3, we can write

$$\lambda_p = \mathcal{A}x_p^m = \lambda(a^{\otimes m})x_p^m + \mathcal{E}x_p^m \ . \tag{3.5}$$

Applying Lemma 1 we get

$$\lambda_p = \lambda(x_p^T a)^m + \mathcal{E}x_p^m = \lambda \cos^m \theta + \mathcal{E}x_p^m \ , \tag{3.6}$$

where θ is the angle between x_p and a. We can use the fact that $|\cos\theta| \leq 1$ together with Lemma 2 and the triangle inequality to obtain the bound

$$|\lambda_p| \leq |\lambda| + \frac{\beta(\mathcal{E})}{m-1} \ . \tag{3.7}$$

We also know that λ_p, as a principal eigenvalue, is a largest-magnitude extremum of the generalized Rayleigh quotient. In particular,

$$|\lambda_p| \geq |\mathcal{A}a^m| \ . \tag{3.8}$$

Now, using Lemmas 1, 2, and 3, we get

$$|\lambda_p| \geq |\lambda + \mathcal{E}a^m| \geq |\lambda| - \frac{\beta(\mathcal{E})}{m-1} \ . \tag{3.9}$$

This establishes the first part of the theorem.

Now, we can combine (3.6) with (3.9) to get

$$\left|\lambda \cos^m \theta + \mathcal{E}x_p^m\right| \geq |\lambda| - \frac{\beta(\mathcal{E})}{m-1} \ . \tag{3.10}$$

Using Lemma 2 we have

$$|\cos^m \theta| \geq 1 - \frac{2\beta(\mathcal{E})}{|\lambda|\,(m-1)} \ . \tag{3.11}$$

□

Theorem 1 means that as $\beta(\mathcal{E})$ approaches zero, then x_p approaches a or $-a$. So, if the noise is small, then the symmetric rank-one approximation of $\mathcal{A}$ corresponding to the principal eigenpair is close to the symmetric rank-one tensor that we seek.

We would like to find the principal eigenpair. However, SS-HOPM will find any eigenvector corresponding to a local maximum of the generalized Rayleigh quotient (or local minimum, under appropriate modifications). The following theorem shows

that if $|\mathcal{A}x^m|$ is sufficiently large and $\beta(\mathcal{A})$ is sufficiently small, then x tells us about a even if it is not a principal eigenvector.

THEOREM 2. *Let $\mathcal{A}$ be defined as in (3.1), and assume, for some x so that $||x|| = 1$, we have*

$$|\mathcal{A}x^m| \geq \epsilon^m + \frac{\beta(\mathcal{E})}{m-1} , \tag{3.12}$$

where $\epsilon > 0$. Then

$$\left|a^T x\right| \geq \epsilon . \tag{3.13}$$

Proof. We have

$$|\mathcal{A}x^m| = \left|(a^T x)^m + \mathcal{E}x^m\right| \tag{3.14}$$

$$\leq \left|a^T x\right|^m + \frac{\beta(\mathcal{E})}{m-1} . \tag{3.15}$$

The proof is by contradiction. Suppose $\left|a^T x\right| < \epsilon$. Then

$$|\mathcal{A}x^m| < \epsilon^m + \frac{\beta(\mathcal{E})}{m-1} . \tag{3.16}$$

But this contradicts our assumption. □

Another interesting question is whether the principal eigenvalue is "well separated" for an approximately rank-one symmetric tensor. Unfortunately, we do not know how to characterize the distribution of the spurious eigenvalues, but we can characterize the distribution of the function of which they are critical points.

THEOREM 3. *Let a be an n-dimensional vector so that $||a|| = 1$. Let x be an n-dimensional vector so that $||x|| = 1$, where x is drawn randomly from the unit sphere. Then*

$$Pr\left(\left|a^T x\right| > \epsilon\right) \leq \frac{1}{n\epsilon^2} . \tag{3.17}$$

As a consequence, if $\mathcal{A}$ is defined by (3.1), then

$$Pr\left(|\mathcal{A}x^m| \geq \epsilon^m + \frac{\beta(\mathcal{E})}{m-1}\right) \leq \frac{1}{n\epsilon^2} . \tag{3.18}$$

Proof. Because of the rotational symmetry of the uniform distribution on the sphere, the distribution of $a^T x$ is identical to $e_i^T x$ for any i, where e_i is a standard basis vector. In particular, $Ee_1^T x = Ea^T x$ and $\mathrm{Var}(e_1^T x) = \mathrm{Var}(a^T x)$. Evidently $Ee_1^T x = 0$ since x is uniform across the unit sphere. So we can write

$$\mathrm{Var}(e_1^T x) = E(e_1^T x)^2 - (Ee_1^T x)^2 = E(e_1^T x)^2 . \tag{3.19}$$

Next, using the symmetry of the uniform distribution, together with the linearity of expectation and the fact that x is unit length, we obtain

$$nE(e_1^T x)^2 = E\sum_{i=1}^{n}(e_i^T x)^2 = E1 = 1 . \tag{3.20}$$

So $\mathrm{Var}(a^T x) = 1/n$. Using Chebyshev's inequality, we can write

$$\Pr\left(\left|a^T x\right| \geq \frac{\kappa}{\sqrt{n}}\right) \leq \frac{1}{\kappa^2} . \tag{3.21}$$

Let $\epsilon = \kappa/\sqrt{n}$, then

$$\Pr\left(\left|a^T x\right| \geq \epsilon\right) \leq \frac{1}{n\epsilon^2} . \tag{3.22}$$

Then (3.18) follows from a direct application of Theorem 2. ▯

Theorem 3 shows that if $\mathcal{A}$ is high-dimensional (n is large), then the generalized Rayleigh quotient is mostly small. Consequently, the principal eigenpair should be a prominent extremum of the generalized Rayleigh quotient.

4. Application of SS-HOPM to approximately-rank-one symmetric tensors. Let us consider the SS-HOPM method applied to the tensor in (3.1). Throughout this section, to simplify discussion, we restrict our attention to $\lambda > 0$. If m is odd, then the eigenvalues come in pairs $\pm\lambda$, one of which is positive, so at least one principal eigenpair is a global maximum of the generalized Rayleigh quotient. If m is even, then Theorem 1 provides that a principal eigenvector x_p must be close to $-a$ or a, which shows us $\lambda_p \approx \lambda > 0$ so it is also a global maximum of the generalized Rayleigh quotient. So, with $\lambda > 0$, we may restrict our attention to negative stable eigenpairs, namely those corresponding to maxima of the generalized Rayleigh quotient, which simplifies discussion of SS-HOPM.

Let us identify a bound on the shift parameter α to guarantee a given negative-stable eigenpair (those corresponding to local maxima) of $\mathcal{A}$, as defined in (3.1), is a stable fixed point of SS-HOPM.

THEOREM 4. *Let $\mathcal{A}$ be defined as in (3.1), and (x_p, λ_p) be a negative-stable eigenpair. Let θ be the angle between x and a. Then x is a stable fixed point for SS-HOPM provided*

$$\frac{-\lambda_p + (m-1)\lambda\left|\sin\theta\cos^{m-2}\theta\right| + \beta(\mathcal{E})}{2} < \alpha . \tag{4.1}$$

Proof. From (2.8), the condition for a stable eigenvector x_p is, for $y \perp x_p$,

$$\left|\frac{(m-1)y^T\mathcal{A}x_p^{m-2}y + \alpha}{\lambda_p + \alpha}\right| < 1 . \tag{4.2}$$

In fact, for negative stable eigenvectors, the expression within the norm is always less than one [4], and we only need to worry about the lower bound

$$-1 < \frac{(m-1)y^T\mathcal{A}x_p^{m-2}y + \alpha}{\lambda_p + \alpha} . \tag{4.3}$$

Applying the definition of $\mathcal{A}$ in (3.1), Lemmas 3 and 1, and the definition of θ, we get

$$-1 < \frac{(m-1)y^T\left(\lambda aa^T\cos^{m-2}\theta + \mathcal{E}x_p^{m-2}\right)y + \alpha}{\lambda_p + \alpha} . \tag{4.4}$$

Using the fact $y^T x_p x_p^T y = 0$, together with the properties of canonical angles between subspaces [9, p. 43], we can write

$$\left|y^T a a^T y\right| = \left|y^T (a a^T - x_p x_p^T) y\right| \tag{4.5}$$
$$\leq \sin\theta \ . \tag{4.6}$$

Together with (2.9), we substitute into (4.4), taking advantage of $\lambda_p + \alpha \geq 0$ (required for convergence), to get

$$-1 < \frac{-(m-1)\lambda \left|\sin\theta \cos^{m-2}\theta\right| - \beta(\mathcal{E}) + \alpha}{\lambda_p + \alpha} \tag{4.7}$$
$$-\lambda_p - \alpha < -(m-1)\lambda \left|\sin\theta \cos^{m-2}\theta\right| - \beta(\mathcal{E}) + \alpha \ , \tag{4.8}$$

and solving for α, we get

$$\frac{-\lambda_p + (m-1)\lambda \left|\sin\theta \cos^{m-2}\theta\right| + \beta(\mathcal{E})}{2} < \alpha \ . \tag{4.9}$$

□

In the limit where $\beta(\mathcal{E})$ is small, we know by Theorem 1 that $\sin\theta$ is small and, using the discussion above to address signs, $\lambda_p \approx \lambda$. So our requirement simplifies to $-\lambda/2 < \alpha$. This bound is much smaller than $\alpha > \hat{\beta}(\mathcal{A})$ provided in [4]. On the other hand, for general eigenvectors where $\sin\theta$ is not small, but $\beta(\mathcal{E})$ is small, our requirement simplifies to $\lambda(m/2-1) < \alpha$. So α in the range $-\lambda/2 < \alpha < \lambda(m/2-1)$, the positive principal eigenvector may be a stable fixed point but spurious eigenvectors may be unstable.

Let us move on to the question of the basin of attraction. To simplify the problem, we consider SS-HOPM applied to an unperturbed rank-one symmetric tensor

$$\mathcal{A} = \lambda \cdot a^{\otimes m} \ . \tag{4.10}$$

It is obvious that the unshifted power method, i.e. SS-HOPM with $\alpha = 0$, converges to a from x in one step provided that $a^T x \neq 0$, because the "range" of the operator $\mathcal{A}x^{m-1}$ consists only of the vector a. We note that if x is chosen randomly, $a^T x \neq 0$ with probability one. When $\alpha \neq 0$, convergence is not obvious, but we can show that under mild conditions, SS-HOPM moves towards the principal eigenvector.

Theorem 5. *Let $\mathcal{A}$ be defined as in (4.10), with $\lambda > 0$. Let x_1 be a vector so that $||x_1|| = 1$, and let $\gamma = a^T x_1$. Assume $\gamma^{m-2} > 0$. Let x_2 be the updated vector under SS-HOPM. Then $\left|a^T x_2\right| > |\gamma|$ provided*

$$\alpha > \frac{-\lambda\gamma^{m-2}}{2} \ . \tag{4.11}$$

Proof. Let us decompose x_1 into its projection onto a and its orthogonal component.

$$x_1 = \gamma a + \delta x_{a\perp} \ . \tag{4.12}$$

Evidently $\gamma^2 + \delta^2 = 1$, and $a^T x_1 = \gamma$. From (2.7), and using Lemma 1, we have

$$x_2 = \frac{\mathcal{A}x_1^{m-1} + \alpha x_1}{||\mathcal{A}x_1^{m-1} + \alpha x_1||} \tag{4.13}$$

$$= \frac{\lambda\gamma^{m-1}a + \alpha\gamma a + \alpha\delta x_{a\perp}}{||\lambda\gamma^{m-1}a + \alpha\gamma a + \alpha\delta x_{a\perp}||} \tag{4.14}$$

$$= \frac{(\lambda\gamma^{m-1} + \alpha\gamma)a + \alpha\delta x_{a\perp}}{\sqrt{(\lambda\gamma^{m-1} + \alpha\gamma)^2 + (\alpha\delta)^2}}\ , \tag{4.15}$$

and so

$$a^T x_2 = \frac{\lambda\gamma^{m-1} + \alpha\gamma}{\sqrt{(\lambda\gamma^{m-1} + \alpha\gamma)^2 + (\alpha\delta)^2}}\ . \tag{4.16}$$

Evidently $\left|a^T x_2\right| > |\gamma|$ is equivalent to

$$\left|\frac{\lambda\gamma^{m-1} + \alpha\gamma}{\sqrt{(\lambda\gamma^{m-1} + \alpha\gamma)^2 + (\alpha\delta)^2}}\right| > |\gamma| \tag{4.17}$$

$$\left|\frac{\lambda\gamma^{m-2} + \alpha}{\sqrt{(\lambda\gamma^{m-1} + \alpha\gamma)^2 + (\alpha\delta)^2}}\right| > 1 \tag{4.18}$$

$$\left|\lambda\gamma^{m-2} + \alpha\right| > \sqrt{(\lambda\gamma^{m-1} + \alpha\gamma)^2 + (\alpha\delta)^2} \tag{4.19}$$

$$(\lambda\gamma^{m-2} + \alpha)^2 > \gamma^2(\lambda\gamma^{m-2} + \alpha)^2 + (\alpha\delta)^2 \tag{4.20}$$

$$(1 - \gamma^2)(\lambda\gamma^{m-2} + \alpha)^2 > (\alpha\delta)^2 \tag{4.21}$$

$$\delta^2(\lambda\gamma^{m-2} + \alpha)^2 > (\alpha\delta)^2 \tag{4.22}$$

$$\delta^2\lambda\gamma^{m-2}(\lambda\gamma^{m-2} + 2\alpha) > 0\ . \tag{4.23}$$

Now, since $\delta^2 > 0$, $\lambda > 0$, and $\gamma^{m-2} > 0$, this is equivalent to

$$\alpha > \frac{-\lambda\gamma^{m-2}}{2}\ . \tag{4.24}$$

□

Let us discuss the requirement $\gamma^{m-2} > 0$. For m even, this is true for all x_1 given $a^T x_1 \neq 0$, and so Theorem 5 provides that SS-HOPM moves ANY input vector towards a with probability one. When m is odd, the property holds for half of the choices of x_1. However, it is easy to check using (4.16) that the sign of γ is preserved under the SS-HOPM update, so repeated applications of SS-HOPM repeatedly improve x_i.

It would be nice to generalize Theorem 5 to the case $\mathcal{E} \neq 0$. However, it cannot hold in the same form because a is not necessarily a stationary point of SS-HOPM in that case. Nonetheless, if the basin of attraction varies smoothly under small perturbation to the original tensor, then we expect the basin of attraction for the principal eigenvector to be large for small $\mathcal{E}$.

We have one more interesting result on the application of SS-HOPM to approximately-rank-one symmetric tensors, but it only holds for even-mode tensors.

THEOREM 6. *Let $\mathcal{A}$ be defined as in (3.1), and assume $\lambda > 0$, m is even, and the shift parameter α for SS-HOPM satisfies $\alpha > \beta(\mathcal{E})$. Then SS-HOPM always increases the generalized Rayleigh quotient and converges to an eigenvector.*

Proof. Define

$$f(x) = \mathcal{A}x^m + (m\alpha/2)(x^T x) \ . \tag{4.25}$$

Notice that the second term of $f(x)$ is constant on the unit sphere, and the SS-HOPM iteration can be written

$$x_{k+1} = \frac{\nabla f(x_k)}{||\nabla f(x_k)||} \ . \tag{4.26}$$

This iteration is known [3, 4] to increase $f(x)$ and converge to an eigenvector provided $\nabla^2 f(x)$ is positive semidefinite symmetric (PSSD). We can write

$$\nabla^2 f(x) = m(m-1)\mathcal{A}x^{m-2} + m\alpha I \tag{4.27}$$
$$= m(m-1)\lambda(a^T x)^{m-2} aa^T + m(m-1)\mathcal{E}x^{m-2} + m\alpha I \ . \tag{4.28}$$

Since $\lambda > 0$ and m is even, the first term is PSSD. So it is sufficient to show that the remaining terms

$$m(m-1)\mathcal{E}x^{m-2} + m\alpha I \ . \tag{4.29}$$

sum to a PSSD matrix. But since the last term is merely a spectral shift, this is assured provided

$$\min_x m\alpha - m(m-1)\rho(\mathcal{E}x^{m-2})) > 0 \ , \tag{4.30}$$

which can be written

$$\alpha > \beta(\mathcal{E}) \ . \tag{4.31}$$

□

We conducted a numerical experiment that illustrates the theorems in this section. We define a tensor $\mathcal{A}$ with $n = 100$ and $m = 4$, and pick $a = (1, 0, 0 \ldots)$. To be able to use n even this large, we need to define $\mathcal{E}$ as a sparse tensor. To generate $\mathcal{E}$, we set $\mathcal{E} = 0$, pick 500 indices at random, and populate those entries with random Gaussian numbers, zero-mean unit-variance. We then permute those indices in all 24 possible ways and copy values to make $\mathcal{E}$ symmetric. Finally, we scale the elements so that $\hat{\beta}(\mathcal{E}) = 0.03$, and so $\hat{\beta}(\mathcal{A}) \approx 3$.

Now, we let α range from -1 to 5, and apply the shifted power method with 10 random starts. Let x be the output of the SS-HOPM, then a success is defined by $\left|a^T x\right| > 0.9$. Figure 4.1 illustrates the success rate as a function of α. To compute α_{min} we combine Theorem 1 and Theorem 4, to get $\alpha_{min} = -0.3365$ for the principal eigenvector and $\alpha_{min} = 1.015$ for the spurious eigenvectors. Evidently the best chance of success for converging to the principal eigenvector is between these two choices of α; the fact that the success rate can be almost 100% is supported by Theorem 5. Choosing $\alpha > \hat{\beta}(\mathcal{A})$, even though it guarantees the SS-HOPM iteration increases the generalized Rayleigh quotient and converges, does not have the best chance of success for recovering the principal eigenvector. We speculate that choosing large α results in more spurious eigenvectors being stable fixed points of the SS-HOPM iteration, resulting in more spurious answers.

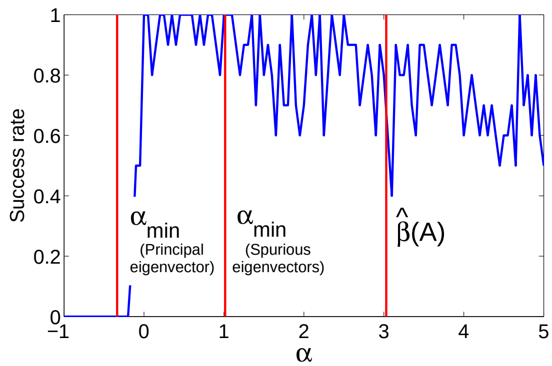

FIG. 4.1. *Success rate for finding the best symmetric rank-one approximation of a symmetric tensor, as a function of shift parameter α. The values for α_{min} come from Theorem 4 for the principal and spurious eigenvectors. Recall $\alpha > \alpha_{min}$ is sufficient but not necessary for stability. The best performance for SS-HOPM on rank-one approximation is when α is about the Theorem 4 threshold for the principal eigenvector but below the threshold for the spurious eigenvectors.*

5. Conclusion. Our perturbative analysis establishes new facts about the structure of approximately-rank-one symmetric tensors, and the application of SS-HOPM to the rank-one approximation problem. We bound the closeness of the best symmetric rank-one approximation, and show that any sufficiently-large eigenpair informs us about the rank-one structure. We show that in high dimensions, most of the generalized Rayleigh quotient, whose critical points correspond to eigenvalues, is close to zero; as a consequence, the principal eigenvalue is prominent. We establish that for rank-one symmetric tensors, under mild conditions, SS-HOPM always moves an input vector towards the principal eigenvector. We also show that the principal eigenvector is a stable fixed point for SS-HOPM under a wide choice of shift parameters, and that SS-HOPM is guaranteed to converge to an eigenvector for a much smaller choice of α in the approximately-rank-one case (for an even number of modes) than the general case. A complete characterization of the basin of attraction for the principal eigenvector remains an open question. Finally, it is hoped that better understanding of the symmetric rank-one problem may lead to better of understanding of more complicated problems such as Independent Components Analysis.

6. Acknowledgements. Thanks to Mark Jacobson, Urmi Holz, Tammy Kolda, Dianne O'Leary, and Panayot Vassilevski for useful observations and guidance.

REFERENCES

[1] LIEVEN DE LATHAUWER, PIERRE COMON, BART DE MOOR, AND JOOS VANDEWALLE, *Higher-order power method—application in independent component analysis*, in Proceedings of the International Symposium on Nonlinear Theory Applications, 1995, pp. 91–96.

[2] CHRISTOPHER HILLAR AND LEK-HENG LIM, *Most tensor problems are NP hard (http://arxiv.org/abs/0911.1393)*.

[3] ELEFTHERIOS KOFIDIS AND PHILLIP A. REGALIA, *On the best rank-1 approximation of higher-order supersymmetric tensors*, SIAM Journal on Matrix Analysis and Applications, 23 (2002), pp. 863–884.

[4] TAMARA G. KOLDA AND JACKSON R. MAYO, *Shifted power method for computing tensor eigenpairs*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1095–1124.

[5] LEK-HENG LIM, *Singular values and eigenvalues of tensors: a variational approach*, in Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor

Adaptive Processing, 2005, pp. 129–132.

[6] V. Olshevsky, ed., *Structured Matrices in Mathematics, Computer Science, and Engineering I*, Contemporary Mathematics, American Mathematical Society, 2001, ch. Tensor approximation and signals processing applications.

[7] Liqun Qi, *Eigenvalues of a real supersymmetric tensor*, Journal of Symbolic Computation, 40 (2005), pp. 1302–1324.

[8] Liqun Qi, Fei Wang, and Yiju Wang, *Z-eigenvalue methods for a global polynomial optimization problem*, Mathematical Programming: Series A, 118 (2009), pp. 301–316.

[9] G. W. Stewart and Ji guang Sun, *Matrix Perturbation Theory*, Academic Press, 1990.

[10] Huahua Wang, Arindam Banerjee, and Daniel Boley, *Modeling time varying covariance matrices in low dimensions*, Tech. Report TR 10-017, Department of Computer Science and Engineering, University of Minnesota, 2010.

[11] Yiju Wang and Liqun Qi, *On the successive supersymmetric rank-1 decomposition of higher-order supersymmetric tensors*, Numerical Linear Algebra with Applications, 14 (2007), pp. 503–519.

[12] Tong Zhang and Gene H. Golub, *Rank-one approximation to high order tensors*, SIAM Journal on Matrix Analysis and Applications, 23 (2001), pp. 534–550.